Learning to Trust

Nooteboom, B.

Publication date:
2005

CentER

Discussion Paper

No. 2005–47

LEARNING TO TRUST

By Bart Nooteboom

September 2003

ISSN 0924-7815

Tilburg University

Learning to trust

Bart Nooteboom
Tilburg University

Paper for a symposium 'La structure cognitive de la confiance', EHESS, Paris, 25-27 September 2003

Abstract

Trust is full of puzzle and paradox. Trust is both rational and emotional. Trust can go beyond calculative self-interest, but has its limits. People may want to trust, while they may also feel threatened by it. If trust is not in place prior to a relationship, on the basis of institutions, prior experience, or reputation, it has to be built up, in specific relations. For that one needs to learn, in the sense of building empathy, and perhaps a certain degree of identification. In an attempt at a better understanding of the puzzles and processes of trust, this chapter applies the perspective of 'embodied cognition', and insights from mental 'framing' and decision heuristics from social psychology.

Key words: learning, trust, institutions

JEL classification: B52, D02, D23, D83

Introduction

As commemorated during the conference on 'La structure cognitive de la confiance', to which this present paper was a contribution, Georg Simmel proposed that trust is a mixture of rationality and feeling: it is based on certain amount of rational assessment, but also entails a leap of faith beyond that. This seems related to the 'paradox of information' associated with trust (Pagden 1988). On the one hand trust requires lack of information: if one were certain about future behaviour, we would no longer speak of trust. On the other hand, trust is based on information, in attributions of motives and competencies to people, based on observed or reported behaviour.

This paper aims to contribute to our understanding of how this mixture of rationality and feeling, assessment and faith, and information and uncertainty may 'work', in the process of the making and breaking of trust. It employs the perspective of 'embedded cognition' (Merleau-Ponty 1942, 1964), which has recently obtained further foundations from recent neuro-science (Edelman 1987, 1992, Damasio 1995, 2003, Lakoff and Johnson 1999). According to this perspective, cognition is rooted in brain and body, which are in turn embedded in their external environment. Here, cognition denotes a broad range of mental activity, including proprioception, perception, sense making, categorization, inference, value judgments, emotions, and feelings.

One core assumption is that people perceive, interpret and evaluate the world according to mental categories that they have developed in interaction with their social and physical environment. This is consistent with the view of 'symbolic interactionism' from sociology (G.H. Mead). It is particularly relevant to trust, which builds up or breaks down in processes of interaction between people. A second assumption is that rationality and emotions, and mind and body, are intertwined (see also Simon 1983, Nussbaum 2001). This goes against the body-mind dualism of Descartes, and is more in sympathy with the thought of Descartes' contemporary Spinoza (Damasio 2003). Events call forth emotions, rooted in the body, which give rise to feelings, which may lead to reflective thought, which may modify emotions and may yield a critical analysis of events. This also is of particular relevance to trust, where emotions play an important role in the way that trust builds up or breaks down in an emotion-laden perception and assessment of threats and risks. For a further understanding of how this works, we will employ insights from social psychology.

In this paper, trust is defined as perceived vulnerability to the actions of others, with the possibility of 'things going wrong', combined with the belief that they will not go (seriously) wrong.
The chapter proceeds as follows. First, it gives a summary of basic notions of trust on which it builds, adopted from the literature (Nooteboom 2002). Second, it gives a further discussion of notions of knowledge and learning used, from the perspective of embodied cognition. Third, it analyses the process of trust building as a process of learning. For a closer analysis of how this works, it next employs insights derived from the theory of framing (Lindenberg 1998, 2003) and decision heuristics from social psychology (Bazerman 1998, Tversky and Kahneman 1983).

Basic features of trust

In line with Nooteboom (2002), trust is here taken as a four-place predicate: the trustor (1) trusts a trustee (2) in on or more aspects of behaviour (3), under certain circumstances (4). Trustees can be individual people, but also collectives, such as organizations, and institutions. The relation between trust in people and trust in organization depends on the position and roles that people have in an organization (Ring and van de Ven 1992, 1994) and on the organization's mode of coordinating behaviour. Concerning aspects of behaviour that one may trust, it is customary to distinguish trust in competence (ability to conform to expectations) and trust in intentions (to perform in good faith according to the best of competence). Competence includes technical and cognitive competence. Trust in intentions requires commitment, i.e. attention to possible mishaps, and absence of opportunism. In the literature, absence of opportunism has been called 'benevolence', 'goodwill' and 'solidarity'. The dependence of trust on circumstances entails that trust is limited: one may trust someone (in competence or intentions), under some conditions but not in others that go beyond competence or resistance to temptations of opportunism.

Concerning the sources of trust, there are psychological causes and rational reasons. Psychological causes include emotions and may entail reflexes or automatic response. Rational reasons entail inference, on the basis of perceived behaviour, of someone's trustworthiness. An important question is how those two sources of trust are related. Can we separate rationality from emotions and feelings? As indicated above, in this article the view is that they cannot be, and that emotions, rationality and feelings are intertwined. According to Damasio (2003), perceptions may trigger emotions, which in turn yield feelings, which may yield thoughts that lead to some constraint on emotions. The question is how this works, in more detail. Assessment of someone's trustworthiness, on the basis of observed or reported behaviour, is limited by uncertainty and bounded rationality, and is mediated by mental heuristics, in perception and attribution of motives and competences of people. Action is based on behavioural routines and their selection, according to decision heuristics. Such heuristics of inference and decision are known from social psychology (Tversky and Kahneman 1983, Bazerman 1998, Smith and Mackie 2000), and will be used in this chapter. Nevertheless, judgements of trustworthiness can be more or less rational, in efforts to avoid 'jumping to conclusions', to be reasonable, and to extend benefit of the doubt to people when trouble occurs. How the psychology of trust may work will be discussed in more detail later in this chapter.

The next question is why people might be trustworthy. Here, we focus on intentional trustworthiness, in particular why people might not act opportunistically. A distinction is made between 'micro' foundations that are specific to a relationship, and 'macro', institution-based foundations that lie in the social environment of a relationship. The distinction between macro and micro sources is also known as the distinction between 'universalistic' or 'generalized' sources versus 'particularistic' sources, made by Deutsch (1973: 55), and between impersonal and personalized sources made by Shapiro (1987) and goes back to the work of Parsons. A distinction is also made between self-interested foundations and foundations that go beyond calculative self-interest. In self-interested foundations, trustworthiness may be based on control or deterrence. The trustor may control opportunities for opportunism ('opportunity control'), or material incentives ('incentive control').

Opportunity control may be based on legal coercion ('macro') or on hierarchical 'fiat', within a relationship ('micro'). Beyond self-interest, and beyond control by the trustor, trustworthiness may be based on socially inculcated values, norms and habits ('macro'), or on personal feelings of empathy or identification, or routinization of conduct in a relationship ('micro'). Empathy entails the ability to understand another's 'way of thinking', without sharing it (having mental models of other people's

mental models), and identification entails that one 'thinks the same way' (having similar mental models). For trust, one needs empathy, but not necessarily identification. One needs to understand 'what makes others tick', without neccessarily 'ticking in the same way'. Empathy is needed to have a sense of the limits of trustworthiness, depending on circumstances. An overview of foundations of trustworthiness is given in Table 1.

Table 1 about here

Note that in Table 1 reputation is included in the self-interested foundations of trustworthiness. Here, one behaves well because bad behaviour would get known in relevant communities, whereby one would forego possibly profitable options for future relationships.

Concerning routinization (see Table 1), Herbert Simon a long time ago showed that routines have survival value due to bounded rationality, in the sense of bounded capacity for reflective thought. Routines allow us to reserve our scarce capacity of 'focal awareness' (Polanyi 1962), in rational, calculative thought, for conditions that are new and demand priority. When things go well for a while in a relationship, one tends to take at least some of it for granted. One may no longer think of opportunities for opportunism open to a partner, or to oneself. On the basis of experience in relations, trustworthiness is assumed until evidence to the contrary emerges. In other words, trust is a 'default'. The possibility of opportunism is relegated to 'subsidiary awareness' (Polanyi 1962). Generally, when something out of the ordinary occurs, our awareness shifts from subsidiary to 'focal' and we look critically at what is going on. As Simon (1983) pointed out, we need emotions of danger and excitement to catapult danger or opportunity into focal awareness. Next, in case of trouble we must control emotions to give the partner the benefit of the doubt, allowing for mishaps, rather than immediately assume the worst (opportunism). In this way, routine behaviour is not necessarily blind, or more accurately: it is not unconditional.

According to the analysis so far, trust may be based on control (coercion, incentives). However, several authors have recognised that trust goes beyond control, in 'goodwill' or 'benevolence' (see e.g. the special issue of Organization Studies on 'Trust and control in organizational relations', 22/2, 2001). As noted by Maguire et. al. (2001: 286), if we do not include the latter, we conflate trust and power. Control or deterrence is part of calculative self-interest, but benevolence is not. Many authors feel that control is foreign to the notion of trust, and that 'genuine' trust is based on other, more social and personal foundations of trustworthiness. Therefore, trust has been defined as the expectation that a partner will not engage in opportunistic behaviour, even in the face of short-term opportunities and incentives (Bradach and Eccles 1984, Chiles and McMackin 1996). To avoid confusion, here the term 'reliance' is used to cover all foundations of trustworthiness, and 'trust' is used for motives that go beyond self-interest.

While trust can go beyond calculative self-interest, in benevolence, it does, and generally should, have its limits. Blind, unconditional trust is generally unwise. Even benevolent people need to guard their self-interest, and not excessively cynical to assume that resistance to temptations to opportunism or betrayal is limited. Managers may be expected to cheat to the extent that their firm is under pressure of survival in competition. An illustration is the ENRON affair. When the overriding survival criterion of a firm is short-term (quarterly) profit, and an economic slump erodes it, the firm may feel irresistible pressure to cheat on the figures. Thus, one should maintain awareness of conditions where trustworthiness may be put under too large a strain.

Yet, as noted before, within limits trust can become routinized and be taken for granted. One does not continually scrutinise behaviour and conditions for opportunities for opportunism, for oneself or one's partner, until they are felt to be excessive. As noted before, in view of uncertainties concerning motives and conditions, trust can only operate as a default: one assumes trustworthiness, within boundaries, until evidence of its failure becomes manifest, and then one adjusts the limits of trust. In other words: one must trust to learn even about trustworthiness. If one only trusted under certainty one would never trust, thereby robbing oneself of the opportunity to learn about trustworthiness and its limits.

One must learn to trust, in finding out how far trustworthiness goes, in different aspects of behaviour. How far does someone's (or a firm's) competence go? Where are the weak spots? How robust is competence under adverse conditions? How strong are pressures of competition, and what slack of resources does a firm have under adversity before it succumbs?
After this summary of the 'basics' of trust, this article focuses on the 'trust process'. The question is on what heuristics of attribution and decision trust and trustworthiness are based, and how this works out in the build-up and break-down of trust.

Knowledge and learning

As indicated in the introduction, this chapter builds on the assumption that knowledge is physically embodied and socially embedded. People perceive, interpret and evaluate the world according to mental categories (or frames or mental models) that they have developed in interaction with their social and physical environment, in 'embodied realism' (Lakoff & Johnson 1999), with the adaptive, selectionist construction of neural nets (Edelman 1987, 1992). The term 'knowledge' here is a broad one, and denotes any mental activity, including perception and value judgements. In other words, we see cognition and emotion (such as fear, suspicion) and body and mind as closely linked (Merleau-Ponty 1964, Simon 1983, Damasio 1995, Nussbaum 2001). The notion that cognition is embedded, arising from interaction with the environment, goes back to Vygotsky (1962) and Piaget (1970, 1974), with their idea that 'intelligence is internalized action'. In sociology, the idea that cognition arises from interaction of people with their (especially social) environment arises, in particular, in the 'symbolic interactionism' proposed by G.H. Mead (1934, 1984).

As a result of differences in physical and cultural environments and individual paths of life that are embodied in cognition, perception, interpretation and evaluation are path-dependent and idiosyncratic to a greater or lesser extent. Different people see and experience the world differently to the extent that they have developed in different physical, social and personal surroundings and have not interacted with each other. In other words, past experience determines absorptive capacity, and there is greater or lesser 'cognitive distance' between people (Nooteboom 1999).

This yields both an opportunity and a problem. Because one cannot 'climb down from one's mind' to assess whether one's knowledge is properly 'hooked on to the world', the variety of perception and understanding offered by other people, on the basis of a variety of experience, is the only source one has for correcting one's errors. Greater distance yields greater novelty value. However, greater distance also makes it more difficult to understand each other and to agree on aims and procedures. If effectiveness of learning by interaction depends on the mathematical product of increasing novelty and decreasing understandability, it has an inverted U-shaped relationship with cognitive distance. This entails a difference between crossing cognitive distance (in understanding people who think differently) and reducing it (thinking more alike). This is the same as the difference between empathy and identification, discussed before. As relationships last longer, cognitive distance is reduced, and identification takes place, especially if the relationship is exclusive, i.e. there are no outside relationships. That is good for trust building but bad for learning. As a result, there is also an inverted U-shaped relation between learning and the duration of a relationship. First, learning increases due to increased understanding, but then learning declines for lack of cognitive distance, in identification. For empirical tests of the hypothesis of optimal cognitive distance, see Wuyts et. al. (2003).

According to Damasio (2003), events, appraised in perception and interpretation, trigger emotions, seen as bodily responses, which may yield automatic response, but also yield feelings that may lead on to critical reflection on the perceived event, its interpretation, consequences, and possible response.

How to begin, to adapt and to end a relationship?

Let us turn to a more detailed analysis of the process of trust development. First, I turn to rational analysis, and psychological processes will be elaborated later. As a transaction relation unfolds in time, one can accumulate more or less reliable information about trustworthiness. And such experience can be communicated in reputation mechanisms. The sociological literature gives extensive instructions how to infer intentional trustworthiness from observed behaviour (Deutsch 1973). Did the partner act not only according to the letter but also to the spirit of the agreement? Did he give timely

warnings about unforeseen changes or problems? Was he open about relevant contingencies, and truthful about his dealings with others who might constitute a threat? Did he defect to more attractive alternatives at the earliest opportunity? Or to use Hirschman's (1970) notions of 'voice' and 'exit': how much voice rather than exit did he exhibit?

In interaction, partners may get to understand each other better, which enables a better judgement of trustworthiness, in 'knowledge based trust'. In ongoing interaction they may first develop insight in each other's cognitive frames, in empathy. This does not entail that they always agree. There may be sharp disagreements, but those are combined with a willingness to express and discuss them more or less openly, in 'voice', extending mutual benefit of the doubt. As a result, conflicts may deepen the relationship rather than breaking it. Next, partners may develop shared cognitive frames, by which they may identify with each other's goals, in 'identification based trust', with understanding or even sympathy for weaknesses and mistakes (McAllister 1995, Lewicki and Bunker 1996).

How, then, does trust develop if there was none before, when there is no basis for ex-ante trust based on earlier experience? The assumption here is that vulnerability cannot be avoided: to achieve its purpose the relationship entails risks of dependence. As indicated earlier (in Table 1), one solution might be to rely on reputation mechanisms. Zucker (1986) suggested that one may infer trustworthiness on the basis of social characteristics, such as upbringing and membership of social groups (such as families, clubs, associations of trade or profession, see also Putnam 2000). Let us suppose, then, that neither reputation nor reliable characteristics are available.

One view is that under those conditions one can only start with control (Lewicki and Bunker 1996), on the basis of contracts, for example, and then shift to trust as knowledge and empathy grow. One problem with that is that in learning and innovation there is likely to be too much uncertainty to specify the conditions of an extended contract, and limited opportunities for monitoring contract execution. Another possibility is develop the relation in a careful balance of mutual dependence, so that there is a threat of retaliation when temptation towards opportunism arises. Another possibility would be to start with small steps, with limited risk, and expand vulnerability as trust grows. One problem with that is that it may take too long. Under present market conditions there is often a need for speed. To reduce risk and to speed up relationship development, one may also profit from the service of specialized intermediaries.

There are a host of different types of intermediaries or go-betweens whose task it is to help judge performance and to provide intermediation or arbitration in conflicts. Shapiro (1987) called these intermediaries 'guardians of trust', Zucker (1986) saw them as part of 'institutions based trust', and Fukuyama (1995) used the term 'intermediate communities'. Many of these serve to develop and police technical or professional standards, with certification systems. There are also roles for go-betweens as consultants in the management of inter-organizational relationships (Nooteboom 2002), in offering arbitration or mediation in conflict, assessing the value of information before it is traded, creating mutual understanding (helping to cross cognitive distance), monitoring information flow as a guard against the spillover of sensitive information, guarding hostages, supporting a reputation system. A further, and perhaps most crucial, role is to act as an intermediary in the building of trust. Trust relations are often entered with partners who are trusted partners of someone you trust. If X trusts Y and Y trusts Z, then X may rationally give trust in Z a chance. X needs to feel that Y is able to judge well and has no intention to lie about his judgment. This can speed up the building of trust between strangers, which might otherwise take too long. Intermediation in the first small and ginger steps of cooperation, to ensure that they are successful, can be very important in the building of a trust relation. The intermediary can perform valuable services in protecting trust when it is still fragile: to eliminate misunderstanding and allay suspicions when errors or mishaps are mistaken as signals of opportunism. He may also help in the timely and least destructive disentanglement of relations. To eliminate misunderstanding, to prevent acrimonious and mutually damaging battles of divorce, a go-between can offer valuable services, to help in 'a voice type of exit'.

Framing

For a deeper analysis, we need to know more about how, in the building and break-down of trust, people make inferences from observed behaviour, and how they act on them. In particular, we want to know how rationality and emotions are combined, and how people go beyond calculative self-interest

and yet refrain from blind, unconditional trust. Here, we employ insights from the theory of framing and from social psychology.

The basic assumptions of framing are the following. People act on the basis of cognitive frames with different motives or goals. At any moment one frame tends to be salient or 'in focus' (Polanyi) while others are in 'subsidiary' awareness. Mental frames are connected with repertoires of action, such as threat, attack, retaliation, defense, surrender, withdrawal, avoidance, approach, offer, receipt, negotiation, etc. In human relations, two basic repertoires of action may be those of 'voice' or 'exit', proposed by Hirschman (1970). Frames, together with situational conditions, trigger behavioural routines that enact the frame. Selection and switches of frames are typically based on emotions, triggered by events in specific contexts of action. Emotions may yield automated, reflexive behaviour, or may lead on to feelings that give rise to rational considerations by which emotions may to some degree be held in check or modified. The importance of emotions here is that they form the trigger of frame switching.

Lindenberg (2003) proposed the following frames :

- 'Guarding one's resources', i.e. focusing on survival or self-preservation
- 'Acting appropriately' (Lindenberg 2003), i.e. according to norms of behaviour (in a community) or shared values (in a specific relationship), gaining social legitimation
- 'Acting as a friend'
- 'Hedonics', i.e. giving in to urges of gratification.

In social psychology, Smith and Mackie (2003) recognized three basic motives of human behaviour: striving for mastery, seeking connections with others, valuing yourself and connected others. Damasio (2003) refers to Spinoza's thought in terms of two basic human drives: towards self perpetuation, which seems similar to Lindenberg's 'guarding one's resources', and towards perfection, which seems similar to what Smith and Mackie recognized as a 'striving for mastery'. Damasio proposed a hierarchy of bodily and mental regulation, as illustrated in figure 1. Here, the 'drives', which Spinoza called 'appetites', of hunger, thirst, sex, etc., seem similar to Lindenberg's 'hedonics'.

Figure 1 about here

Here, for maximum simplicity, but in broad agreement with the typologies indicated above, I assume two basic sets or families of frames: self-directed (including concern for survival, resources, gratification) and other-directed (acting appropriately, as a friend, connections with others, social legitimation). Note that this brings us close to the classification of sources of (intentional) reliability in Table 1.

Stability of relations depends on frame stability, which depends on how salient a frame is, which depends on how strongly it is held, on what frames are subsidiary and on the extent to which they are complements or substitutes to the salient frame (Lindenberg 2003). If, for example, the salient frame is to act in regard for others, and the frame of self-interest is subsidiary, they complement each other when self-interest is served by collaboration with others. When self-interest is threatened, beyond some tolerance level, the stability of an other-directed frame is precarious. This is how I reconstruct the limits of trust in psychological terms.

Decision heuristics

The question now is what heuristics are used in these processes of the formation, selection and enactment of frames. Here, I turn to decision heuristics proposed in social psychology by Bazerman (1998):

- Representativeness heuristic: the likelihood of an event is assessed by its similarity to stereotypes of similar occurrences.
- Availability heuristic: people assess the probability and likely causes of an event by the degree to which instances of it are 'readily available' in memory, i.e. are vivid, laden with emotion, familiar, recent and recognizable. Less available events and causes are neglected.

- Anchoring and adjustment. Judgement is based on some initial or base value ('anchor') from previous experience or social comparison, plus incremental adjustment from that value. People have been shown to stay close even to random anchors that bear no systematic relation to the issue at hand. First impressions can influence the development of a relation for a long time.

These heuristics serve to give more substance to the notion of absorptive capacity, i.e. the ability to perceive and interpret phenomena, and to the claim, made in embodied cognition, that rationality and emotions are intertwined. The heuristics are not rational in a calculative sense (*calculative rationality*). Indeed, they serve to show how bounded rationality works. However, they are 'adaptively rational' in the sense of contributing to survival under uncertainty and bounded rationality, and the need, in many situations, to decide and act quickly (*adaptive rationality*). Nevertheless, they can lead to error, as will be discussed. In the elaboration of these heuristics I present what I make of them, from the perspective of embodied cognition and framing theory, and this may deviate from established practice in social psychology.

Prospect theory (Tversky and Kahneman 1983) has demonstrated that people are not risk-neutral, and tend to be risk-taking when a decision is framed in terms of loss, and risk-averse when it is framed in terms of gain. This entails that people will accept a greater risk of conflict when they stand to incur a loss than when they stand to obtain a gain. As a result, the frame of guarding resources, or going for material self-interest, splits up into a frame of loss and a frame of gain. Related to this effect is the *endowment effect*: people often demand more money to sell what they have than they would be prepared to pay to get it. In the first case one wants to cover for loss.

Yet another psychological mechanism is that in violation of rational behaviour sunk costs, such as sacrifices made in a relationship, are not seen as bygones that should be ignored in an assessment of future costs and benefits. They are seen as sacrifices that would be seen as in vain if one pulls out after having incurred them. This yields what is known as *non-rational escalation of commitment*. It is associated with *cognitive dissonance*: cutting one's losses and pulling out would entail an admission of failure, of having made a bad decision in the past. Deutsch (1973) gave the example of the US finding it increasingly difficult to pull out of Viet Nam as the number of killed soldiers accumulated. The phenomenon is confirmed in empirical research, which shows that when the decision is to be made by someone not involved in the initial commitment, or when the threat of an admission of failure is removed, the rational decision to pull out is made. Again, one cannot say that this mechanism is always bad, because it also demonstrates perseverance in the face of setbacks, which can be a good thing, and is in fact a trait of many a successful innovating entrepreneur. This phenomenon can be connected with the effect of a loss frame versus a gain frame. The person, or group, that made the initial decision experiences a loss frame, with the inclination to accept further risk in order to prevent acceptance of the loss. The decision maker who enters fresh experiences a gain frame, to make a decision that will offer profit or prevent further loss in the future, regardless of past sunk costs.

Evolutionary psychologists claim that certain psychological features or mechanisms are 'in our genes' as a result of evolution (Barkow et. al. 1992). They emerged as features that gave selective or reproductive advantage, over the millions of years that the human species evolved in hunter-gatherer societies. For example, survival required the basic ability to identify objects and movement, to categorize natural kinds (plants, animals), distinguish the animate from the inanimate, and natural kinds from artefacts (Tooby and Cosmides 1992: 71). On top of that, it requires the ability to recognise objects, judge speed and distance, to avoid predators and to catch prey (Tooby and Cosmides 1992: 110). Survival also requires mother-infant emotion communication signals (Tooby and Cosmides 1992: 39). Such instincts contribute to the heuristics of availability and representatives, in our framing of the world.

These heuristics and principles from social and evolutionary psychology are consistent with the perspective of embodied cognition, and indeed serve to elucidate and extend it, in their integration with emotions, their pragmatic role in survival under conditions of radical uncertainty, and their embedding in processes of practical action.

Application to trust

In my interpretation, the *representativeness heuristic* constitutes our repertoire for categorization, i.e. what we *can* attend to, in our absorptive capacity. There is a connection with the role of 'prototypes' or 'exemplars' in language and categorization (Rosch 1978, Nooteboom 2000). Since definitions can seldom offer necessary and sufficient conditions for categorization, and meaning is context-dependent and open-ended, allowing for variation and change, we need prototypes. Prototypes are salient exemplars of a class that guide categorization by assessing similarity to the prototype. We try to fit or assimilate observed behaviour into prototypes in our cognitive repertoire, and when we recognise some features as fitting, we tend to attribute remaining, unobserved attributes that belong to the prototype. The mechanism of attributing unobserved characteristics upon recognition of observed ones enables fast pattern recognition, which is conducive to survival. The downside of the representativeness heuristic is that it also yields prejudice, in the premature, erroneous application of stereotypes, in mistaken attributions. In the present framework, the representativeness heuristic regulates the *cognitive content* of frames, in terms of the categories that they employ. In the context of trust, I see the representativeness heuristic as providing benchmarks, in the form of prototypes, for efficient, fast identification of trustworthy and untrustworthy behaviour, and guidelines or exemplars for trustworthy behaviour. In organizations, such prototypes for trust are often part of organizational culture. An example is the following. In her analysis of 'trust and trouble' in organizations, Six (2004) found the precept, in one organization, that in case of trouble, 'people should not complain *about* people but *to* them'. Here, the organizational ethic is one of voice: when trouble arises, be open about it and try to work it out together.

The *availability heuristic*, in my interpretation, regulates what we *actually* attend to, by filtering impressions, in emotions that contribute to the selection of frames. If we did not apply such filters, our consciousness would likely be overloaded. We cannot afford to pay attention to everything that is presented to our senses, and we need to select what appears to be salient and urging attention. Much of our conduct is based on routines that are relegated to subsidiary awareness. Then, as discussed earlier, we need emotions to catapult attention back into focal awareness when a threat or new opportunity emerges. Emotions tend to be stronger when personal desires or interests are at stake than in more abstract, impersonal motives. Thus, frames of hedonism and 'guarding resources' tend to acquire salience more easily than a normative frame (Lindenberg 2003). However, when the desire to 'act appropriately' is based on emotions of friendship or kinship, it will have greater 'availability'.

As noted before, trust may become routinised, when a relationship has been going well for a while. Trustworthiness is taken for granted, until something exceptional arises, in observed behaviour or conditions that might yield a temptation or pressure that a partner may not be able to resist. This evokes feelings of fear, suspicion or indignation that break routinized trust open to critical scrutiny. Earlier, we noted the role of identification in trust, on the basis of shared categories concerning motives and conditions of behaviour. Here, availability is based on familiarity. It affects both one's own trustworthiness, in the willingness to make sacrifices for others, and one's trust, in the tolerance of behaviour that deviates from expectations. One will more easily help someone when one can identify with his need. One can more easily forgive someone's breach of trust or reliance when one can identify with the lack of competence or the motive that caused it. One can more easily accept the blame for oneself. Since one can identify with him, one may sympathize with his action, seeing, perhaps, that his action was in fact a just response to one's own previous actions. One might have reacted in the same fashion as he did.

While it is adaptively rational, the availability heuristic yields several problems. One is that anger may overshoot its function of calling attention, and propels impulsive defensive or retaliatory conduct. Another problem is that attention is called only by extreme, emotional appeals, so that more subdued, nuanced appeals and weak signals tend to be ignored. The availability heuristic yields useful emotional triggers to create focal awareness of possible risk in relations, but next one should be careful not to jump to conclusions, and to exercise benefit of the doubt wherever possible. This is the case, in particular, because when 'things go wrong' in a relationship, there may be a multitude of possible causes: an accident, lack of resources, lack of competence and opportunism. One may jump to assuming the worst, opportunism, while that conclusion is not justified.

Anchoring and adjustment indicates that once we select a frame, with corresponding behavioural routines, we do not easily drop it. Deutsch (1973) also argued that beginnings are important and may be difficult to turn around. He suggested that there is circular causation between characteristics of participants and the results of interaction, in his 'crude law of social relations': 'The characteristic

processes and effects elicited by a given type of social relationship (cooperative or competitive) tend also to elicit that type of social relationship'. Under uncertainty cognition does need such an anchor. Studies of learning and adjustment have shown that hasty and large departures from existing practices can yield chaotic behaviour (March 1991, Lounamaa and March 1987). It is adaptively rational to experiment with small, proximate change, to accumulate motivation for more radical change, as well as insights into directions and elements of change that are likely to be viable and worth the upheaval (Nooteboom 2000). However, this heuristic also entails risk of inertia: the inability to make needed drastic change in the face of a crisis.

This heuristic is related to the notion of trust as a default, and routinized trust. The notions of default and routine entail that one sticks to an established form of behaviour until new evidence compels its change.

As already noted, the heuristic of anchoring and adjustment entails that beginnings are often difficult to turn around. This relates back to the discussion of how to start a relationship when at the beginning there is no basis for trust. If one takes the advice from Lewicki and Bunker to start on the basis of control, this may be seen as a sign of distrust, which is not only likely to evoke a similar response, with the risk of a vicious circle of control upon control that constrains the scope for flexibility, improvization and innovation in the relationship, but may also be difficult to turn into trust as the relationship develops.

These heuristics appear to complement each other. Representativeness determines how one can interpret behaviour, availability determines which interpretation is triggered, and anchoring stabilises chosen behavioural routines. Of the three, availability is the most laden with emotion, and representativeness is the most concerned with learning. Availability causes frame switches, and anchoring and adjustment serves to constrain such switches. The first is needed for guarding survival, the latter is needed to make framing less unstable and precarious.

The notion of a *loss versus a gain frame*, from prospect theory, has important implications for the stability of relationships (Nooteboom 2004). It was noted earlier that one way to start a relationship without prior trust is to develop it in a careful balance of mutual dependence. Then, however, the problem often is that the balance is broken by a change of conditions, such as a change of technology or market whereby existing competences lose relevance, in different degrees for different partners, or the emergence in the arena of a new player who yields a more attractive alternative for one of the partners. This partner may then want to exit while the other partner wants to hold on to him. Then, the first partner is in a gain frame (getting more out of an alternative relationship), while the other is in a loss frame (losing a valuable partner). According to the theory, the latter would go to greater extremes, in emotional and even self-destructive actions (slander, hopeless litigation), to retain the partner, than the other partner would engage in to realise his exit.

The phenomenon of *escalation of commitment* also contributes to the stability of a relationship. After a relationship has cost an accumulation of sacrifice, it may become increasingly difficult to exit from it, since that would suggest that past sacrifices were wasted. Such partnership may quickly unravel under new management that is not compromised by past sacrifices.

In sum, there appear to be three mechanisms for the stabilization of relationships: anchoring, loss versus gain, and escalation of commitment.

Evolutionary psychology claims that in the hunter-gatherer societies in which man evolved, the variance of yields, in gathering edible plants, roots, nuts, etc., and even more in hunting, with the impossibility to instantly consume large game, together with problems of durable storage, entails an evolutionary advantage of *reciprocity*, in the willingness to surrender part of one's yield to others in need, in the expectation to receive from them when they are successful (Cosmides and Tooby 1992: 212).[1] This is enhanced by the supposed ability to assess such willingness among others, in a 'cheater detection mechanism', and to signal a credible, often emotion-laden threat to sanction lack of reciprocity (Frank 1988). This purported natural inclination towards reciprocity would contribute to a solution of the problem, discussed before, of how to start a relationship without trust based on prior experience, while allowing trust to develop. Presumably, the principle works only under conditions of

[1] The argument requires that group selection is viable. In spite of earlier arguments that group selection is dominated by individual selection, according to later arguments it can be viable. That is the case if opportunistic intruders into a reciprocity-based society are pre-empted before they get a chance to develop reproductive advantage (Ridley 1997).

stability of communities, where recipients of goodwill will 'stay around' to engage in reciprocity. A similar condition was discussed for a reputation mechanism. Reciprocity and reputation are mutually reinforcing.

The heuristics have implications not only for starting and adapting a relationship, but also for ending one. The ending of relationships is at least as important as the issue of starting one, in view of the fact that relationships may last too long, yielding too much identification, resulting in lack of cognitive variety for further innovation. Nooteboom (1999, 2004) proposed that there is a hostile and a collaborative mode of exit from a relationship. In the hostile mode, one would try to prepare one's exit, no longer engaging in relation-specific investments, retrieving hostages, and building up investments in a new relationship, as surreptitiously as one can, and then spring a surprise exit on the partner, who is left with the unforeseen damage of worthless specific investments and a discontinuity in production. In the collaborative mode, in a 'voice mode of exit', one would announce one's intentions towards exit ahead of time, help the partner to disentangle the relationship with minimum damage, and help to find a replacement. The advantage of the latter mode would be that the partner is less in a loss frame, and hence less likely to take radical actions of binding and retaliation. A risk, however, is that the partner would have more time and opportunity to take such actions. Here, the anchoring and adjustment heuristic also kicks in. If the relationship had been collaborative, based on benevolence, that would set the norm, and a hostile mode of exit would constitute a greater shock than in a more calculative relationship, with greater risks of extreme retaliatory behaviour.
In sum, the decision heuristics and other phenomena from social and evolutionary psychology are highly relevant to the development of trust, because they affect attribution of characteristics, expectations of trustworthiness, and choice of action.

Conclusions

A central claim of this paper is that for an understanding of how trust is built up and broken down, and how trust can go beyond calculative self-interest without becoming blind or unconditional, we may profit from the application of framing theory and decision heuristics derived from social psychology. They are consistent with the perspective of embodied cognition, and yield an elaboration of it. In particular, they elaborate the idea that rationality and emotions are interwined, and that psychological mechanisms arise from conduct that is selected for survival. In this article, an attempt was made to combine framing theory and decision heuristics. A frame is an intentional stance, guiding perception and action, which focuses on a limited number of a range of possible other-directed or self-interested goals or motives. At any moment one frame tends to be salient while others are subsidiary, and perceptions of events may cause switches.

This helps to understand the limits of trust. Decision heuristics help to explain how this happens. The following heuristics are discussed:

- The representativeness heuristic
- The availability heuristic
- The anchoring and adjustment heuristic
- The phenomenon of escalation of commitment
- The phenomenon of cognitive dissonance
- The notion of a loss frame versus a gain frame
- The notion of an instinct for reciprocity, combined with a cheater detection mechanism

The representativeness heuristic was interpreted as yielding a repertoire of possible frames and behavioural routines, constituting a cognitive repertoire, in the form of prototypes, in terms of which judgements of (un)trustworthiness are made and translated into action. The availability heuristic was interpreted as determining which frames are salient and which are subsidiary, and how switching between them occurs. This is governed by emotions and by familiarity. Emotions serve to break behavioural routines open for scrutiny. On the other hand, the heuristic of anchoring and adjustment, escalation of commitment, and the notion of a loss versus a gain frame serve to stabilise behaviour. Anchoring and adjustment clarify why it may be difficult to turn suspicion around into trust and why it is difficult to take a hostile, abrupt approach to exit if the relationship was based on benevolence and voice. The notion of an instinct towards reciprocity helps to start a relationship in which there is no prior trust based on experience.

Here also we find paradoxes. These heuristics and phenomena are adaptively rational, given bounded rationality and the pressure to survive under conditions of radical uncertainty and the need for speedy response. However, they can yield irrational actions, in prejudice, blindness, injustice, inertia, lack of adaptiveness under crises, and impulsive behaviour. They help to enable and support the building of trust, but may also prevent it, and may cause unjustified breakdowns of trust.

A recurrent theme in the process of trust is openness, in a voice based approach. The importance of that was argued earlier by Zand (1972). Since there may be many causes of disappointed expectations, it is important to be open about problems that one may cause to a partner. Silence until the problem becomes manifest is likely to be interpreted as a sign of the worst possibility of opportunistic behaviour. An imposition of control and requisite monitoring has adverse effects, indicated before, of a vicious circle of control that blocks scope for learning and innovation, and starts a relation on a footing of mistrust that may be difficult to turn around. If, on the other hand, one voluntarily accepts or even invites some control by the partner, and opens up to monitoring, this greatly facilitates trust building. This indicates the importance of voluntary transparency for trust relationships. A timely announcement of intentions to exit may greatly help to disentangle relationships without too much damage. If one shies away from exhibiting one's fears of relational risk, one thereby robs the partner of the opportunity to reduce that risk.

While some progress may have been made in this article, in the attempt at a coherent account of the trust process, in a combination of framing theory and decision heuristics, there is need for further research, analysing in more detail how the different psychological mechanisms mesh and compensate or conflict, and how this depends on details of interaction and on social contexts, such as organizational culture and institutional environments. There is much scope and need for further empirical research, to test and revise the many conjectures developed in this article. In cognitive science, research in embodied cognition may investigate the neural correlates or underpinnings of the psychological mechanisms discussed in this article.

References

Barkow, J., L. Cosmides and J. Tooby (1992). The adapted mind: Evolutionary psychology and the generation of culture, Oxford: Oxford University Press.

Bazerman, M. (1998). Judgment in managerial decision making, New York: Wiley.

Bradach, J.L. and R.G. Eccles (1984). 'Markets versus hierarchies: From ideal types to plural forms', in W.R. Scott (ed.), Annual Review of Sociology, 15: 97-118.

Chiles, T.H. and J.F. McMackin (1996). 'Integrating variable risk preferences, trust and transacton cost economics', Academy of Management Review, 21/7: 73-99.

Cosmides, L. and J. Tooby (1992). 'Cognitive adaptations for social exchange', in H. Barkow, L. Cosmides and J. Tooby, The adapted mind, Oxford: Oxford University Press: 163-228.

Damasio, A. R. (1995). Descartes' error: Emotion, reason and the human brain, London: Picador.

----- (2003). Looking for Spinoza, Orlando, Florida: Harcourt.

Deutsch, M. (1973). The resolution of conflict: constructive and destructive processes, New Haven: Yale University Press.

Edelman, G. M. (1987). Neural Darwinism; the theory of neuronal group selection, New York: Basic Books.

----- (1992). Bright air, brilliant fire; On the matter of mind, London: Penguin.

Frank, R.H. (1988). Emotions within reason, the strategic role of the emotions, New York: W.W. Norton.

Fukuyama, F. (1995). Trust, the social virtues and the creation of prosperity, New York: Free Press.

Hirschman, A.O. (1970). Exit, voice and loyalty: Responses to decline in firms, organisations and states, Cambridge MA: Harvard University Press

Lakoff, G. and M. Johnson (1980). Metaphors we live by, Chicago: University of Chicago Press.

----- (1999). Philosophy in the flesh, New York: Basic Books.

Lewicki, R.J. and B.B. Bunker (1996). 'Developing and maintaining trust in work relationships', in R.M. Kramer T.R. Tyler eds , Trust in organizations: Frontiers of theory research, Thous Oaks: Sage Publications: 114-139.

Lindenberg, S. (1998). 'Solidarity: its microfoundations and macro-dependence: A framing approach', in P. Doreian and T.J. Farro (eds.), The problem of Solidarity: Theories and models, Amsterdam: Gordon and Breach: 61-112.

----- (2003). 'Governance from a framing point of view: The employment relationship and relational signalling', in B. Nooteboom and F. Six, The trust process in organizations, Cheltenham (UK): Edward Elgar: 37-57.

Lounamaa, P. H. and J. G. March (1987). 'Adaptive coordination of a learning team', Management Science, 33: 107 - 23.

Maguire, S., N. Philips and C. Hardy (2001). 'When "silence=death", keep talking: Trust, control and the discursive construction of identity in the Canadian HIV/AIDS treatment domain', Organization Studies, 22/2: 285-310.

March, J. (1991). 'Exploration and exploitation in organizational learning', Organization Science, 2/1, 101-123.

McAllister, D.J. (1995). 'Affect- and cognition based trust as foundations for interpersonal cooperation in organizations', Academy of Management Journal, 38/1: 24-59.

Mead, G. H. (1934). Mind, self and society; from the standpoint of a social behaviorist, Chicago: Chicago University Press.

----- (1982). The individual and the social self, unpublished work of G.H. Mead, edited by D.L. Miller, Chicago: University of Chicago Press.

Merleau - Ponty, M. (1942). La structure du comportement, Paris: Presses Universitaires de France.

----- (1964). Le visible et l'invisible, Paris: Gallimard.

Nooteboom, B. (1999). Inter-firm alliances: Analysis and design, London: Routledge.

----- (2000). Learning and innovation in organizations and economies, Oxford: Oxford University Press.

----- (2002). Trust: Forms, foundations, functions, failures and figures, Cheltenham (UK): Edward Elgar.

----- (2004). Inter-firm collaboration, learning and networks: An integrated approach, London: Routledge.

Nussbaum, M.C. (2001). Upheavals of thought, The intelligence of emotions, Cambridge UK: Cambridge University Press.

Pagden, A. (1988). 'The destruction of trust and its economic consequences in the case of eighteenth-century Naples', in D. Gambetta (ed.), Trust, the making and breaking of cooperative relations, Oxford: Blackwell: 127-141.

Piaget, J. (1970). Psychologie et epistémologie, Paris: Denoël.

----- (1974). Introduction a l'épistémologie génétique, I and II, Paris: Presses Universitaires de France.

Polanyi, M. (1962). Personal knowledge, London: Routledge.

Putnam, R.D. (2000). Bowling alone; the collapse and revival of American community, New York: Simon and Schuster.

Ridley M. (1997). The origins of virtue, Viking Press.

Ring, P. Smith and A. van de Ven (1992). 'Structuring cooperative relationships between organizations', Strategic Management Journal, 13: 483-498.

----- (1994). 'Developmental processes of cooperative interorganizational relationships', Academy of Management Review, 19/1: 90 - 118.

Rosch, E. (1978). 'Principles of categorization', in E. Rosch and B. B. Lloyd (eds.), Cognition and categorization, Hillsdale, NJ: Erlbaum: 27-48.

Shapiro, S.P. (1987). 'The social control of impersonal trust', American Journal of Sociology, 93: 623-658.

Simon, H. A. (1983). Reason in human affairs, Oxford: Basil Blackwell.

Six, F. (2004). Trust and trouble in organizations, PhD thesis, Rotterdam School of Management, Erasmus University Rotterdam, forthcoming.

Smith, E.R. and D.M. Mackie (2000). Social psychology, Philadelphia: Taylor and Francis.

Tooby, j. and l. Cosmides (1992). 'The psychological foundations of culture' in J.H. Barkow, L. Cosmides and J. Tooby, The adapted mind, Oxford: Oxford University Press: 19-136.

Tversky, A. and D. Kahneman (1983). 'Probability, representativeness, and the conjunction fallacy', Psychological Review, 90/4: 293-315.

Vygotsky (1962). Thought and language, edited and translated by E. Hanfmann and G. Varkar, Cambridge MA: MIT Press.

Wuyts, S., M.G. Colombo, S. Dutta and B. Nooteboom (2003). Empirical tests of optimal cognitive distance, Research Report, Erasmus Institute for Research in Management, Erasmus University Rotterdam.

Zand, D.E. (1972). 'Trust and managerial problem solving', Administrative Science Quarterly, 17/2: 229 - 239.

Zucker, L.G. (1986). 'Production of trust: Institutional sources of economic structure', in Barry, Staw & Cummings, Research in organizational behaviour, vol. 8: 53-111.

Table 1 Sources of (intentional) reliability

	macro; universalistic	**micro**; particularistic, relation-specific
self-interest		
opportunity control	contracts, legal enforcement	hierarchy, managerial 'fiat',
incentive control	reputation	dependence: unique partner value, switching costs, hostages
altruism benevolence	values, social norms of proper conduct,, moral obligation, sense of duty, bonds of kinship	empathy, routinization, identification, affect, friendship

source: adapted from Nooteboom (2002).

Figure 1: hierarchy of bodily and mental regulation

Feelings

Emotions:

social emotions: sympathy, embarrassment, shame, guilt, pride, jealousy, envy, gratitude, admiration, indignation, contempt, …

primary emotions: fear, anger, disgust, surprise, sadness, happiness, …

background emotions (state of being): enthusiasm, edginess, excitement, tranquillity, ..

Drives (appetites):

Hunger, thirst, curiosity and exploration, play, sex, ..

Pain and pleasure:

Yielding reflexes of withdrawal, attraction, protection, expression of alarm

---

Immune response

Basic reflexes:

Startle reflex, ..

Metabolic regulation:

Endocrine/hormonal secretions, muscular contraction (heart, digestion, ..), heart rate, breathing, blood pressure, storage and deployment of proteins and carbohydrates, …

Adopted from Damasio (2003).